Artificial Intelligence for territorial interpretation: from image clustering to perceptual mapping

Fabio Bianconi[1], Marco Filippucci[1], Chiara Mommi[1]

[1] Department of Civil and Environmental Engineering, University of Perugia, 06125 Perugia, Italy.
fabio.bianconi@unipg.it, marco.filippucci@unipg.it, chiara.mommi@dottorandi.unipg.it

Keywords: Artificial Intelligence, Image Clustering, Landscape Interpretation, Multi-Agent Evaluation, Cognitive Mapping, Data-Driven Representation

Abstract

The research investigates artificial intelligence as a device for the automatic interpretation of landscape, reframing representation not as a neutral reproduction but as a cognitive operation in which perception, description, and evaluation converge. Moving from the assumption that landscape is not an objective given but a culturally and perceptually constructed form, the study proposes a fully data-driven methodology based on geolocated images. Through a systematic grid sampling, street-level panoramic views are collected and processed within a multimodal pipeline integrating visual analysis, language models, and multi-agent evaluation. Images are first translated into textual descriptions and semantically clustered, allowing territorial classes to emerge from the data rather than from predefined taxonomies. In parallel, a simulated cognitive framework, structured through four agent profiles, produces evaluative scores and textual judgments, later analysed through sentiment detection. The integration of these layers generates a georeferenced dataset from which a perceptual cartography of the territory is constructed. Applied to the urban context of San Giustino (Italy), the method reveals a continuous gradient from dense urban cores to rural landscapes, while exposing differentiated perceptual readings across observer profiles. Within this framework, artificial intelligence does not replace human interpretation; it operates as an epistemic extension, transforming the landscape into a distributed field of comparable perceptions, where representation becomes a computable form of shared knowledge.

1. Introduction

The landscape is not an objective reality, but a way of interpreting what is seen (Bianconi & Filippucci, 2019; D. E. Cosgrove, 2008; Mitchell, 2002). Its form originates in the gaze that observes it and in the image that represents it within the mind. In this dual dimension, both perceptual and figurative, lies its cognitive nature: the landscape operates as a device that translates objective experience into a subjective visual structure, making visible the relationships between matter, time, and memory, and thus between place and person (Berque, 1995; Bianconi & Filippucci, 2021).

Given its intrinsic complexity, the landscape has historically been interpreted through two main approaches. The geographer observes it from above, adopting a zenithal perspective that measures boundaries and describes spatial structures; the artist experiences it from within, capturing its perceptual depth through immersive and perspectival representation. Both approaches are essential, yet partial: one abstracts and rationalises, the other interprets and immerses. Neither, on its own, is sufficient to grasp the full cognitive and experiential dimension of the landscape (D. Cosgrove, 1984; Lynch, 1960; Corner, 1999).

Every territory is the result of a continuous stratification of natural processes and human actions. Geological formations, climatic conditions, settlement patterns, and infrastructures intertwine over time, generating a complex spatial palimpsest in which traces of different historical phases coexist. Through this process, the territory becomes landscape: not merely a physical configuration, but a cultural and perceptual construct, a form of visible collective memory (Corboz, 1985; Lynch, 1960; Antrop, 2005; Wu, 2013). Representation, therefore, cannot be understood as a neutral reproduction of reality. To represent a landscape means to select, interpret, and reorganise its elements, revealing latent structures and underlying relationships through cognitive and graphic processes (de Rubertis, 1994; Ugo, 1989). Drawing, in this context, assumes a central role (Carpo, 2017). It is not only a tool of communication but a form of thought, a cognitive operation through which perception is transformed into structured knowledge. Through signs, schemes, and visual relationships, drawing organises complexity and produces meaning, mediating between sensory experience and conceptual understanding (de Rubertis, 1994). The act of representation thus unfolds within a constant tension between objectivity and subjectivity: on one side, the need for measurable, shareable, and verifiable data; on the other, the interpretive role of the observer, whose cultural background and perceptual sensitivity inevitably shape the image produced (Pallasmaa, 1994). Today, this cognitive operation of selecting, interpreting, and reorganising reality is no longer confined to human drawing, but finds a powerful new medium in computational algorithms.

In recent years, digital technologies, and in particular artificial intelligence (AI), have significantly expanded the possibilities for landscape interpretation (Mai et al., 2024; Xing et al., 2025). Advances in computer vision and deep learning have enabled the large-scale analysis of visual environments, allowing spatial, semantic, and perceptual information to be extracted directly from images (Ma et al., 2019; Peng et al., 2025; Zhou et al., 2018). More recently, multimodal approaches have further extended these capabilities by integrating visual and linguistic processing, enabling systems not only to recognize scenes but also to describe and interpret them through natural language (Bianconi et al., 2024; Zhang et al., 2025). Within this framework, the landscape can be understood as a distributed dataset of visual signals, in which each image encodes latent information on spatial form, perceptual quality, and environmental relationships (Bommasani, 2021; Carpo, 2023, Wei et al.,2022).

Artificial intelligence does not replace human vision; rather, it extends it. It introduces a new level of mediation between perception and representation, enabling the integration of quantitative analysis and qualitative interpretation within a unified cognitive process (Floridi, 2019; Manovich, 2020). Through automated description, semantic clustering, and simulated evaluation, AI systems can construct distributed and comparable readings of the territory, transforming individual perception into structured and shareable data. The image thus becomes not only an aesthetic object but also a unit of geographic knowledge, a medium through which representation, interpretation, and evaluation converge (Mai et al., 2024; Zhang et al., 2025)

Within this paradigm shift, the present research explores the use of artificial intelligence as a tool for automatic territorial interpretation. The study proposes a data-driven methodology based on geolocated images, combining visual analysis, linguistic processing, and multi-agent evaluation to construct a perceptual map of the landscape. The underlying assumption is that the landscape can be read from the ground up, starting from its visual traces, without relying on predefined categories. In this sense, AI acts as a distributed observer, traversing the territory through images and recognising patterns, forms, and perceptual qualities that emerge directly from the data.

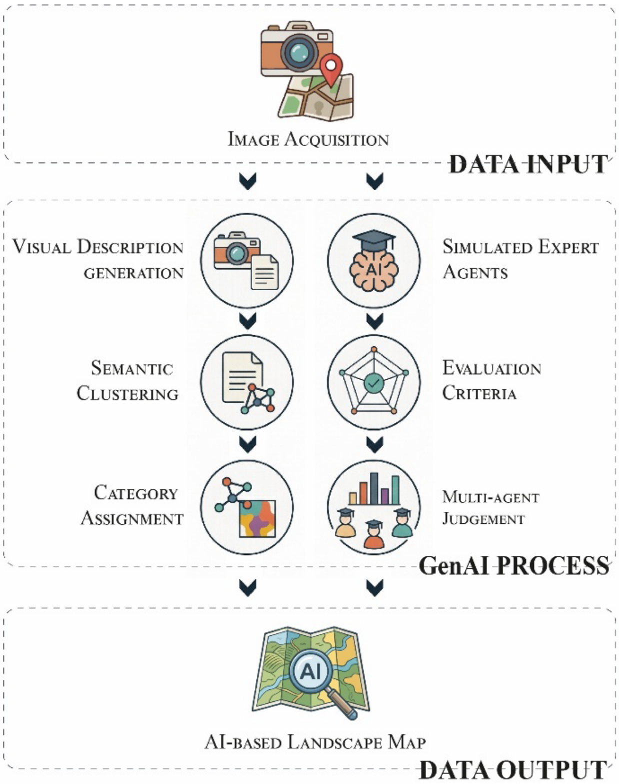

Figure 1. Data-driven workflow combining visual, linguistic, and cognitive processes. Starting from geolocated images acquired via © Google Maps, the pipeline integrates automatic description, semantic clustering, and multi-agent evaluation to generate an AI-based landscape map with classified, sentiment, and perceptual data for each point.

2. Methodology

The proposed methodology is based on a fully data-driven approach that integrates Large Language Models for the interpretation of the territory through geolocated images (Figure 1).

The first phase concerns the definition of the spatial domain and the construction of the visual dataset. The study area is defined starting from a georeferenced centroid expressed in coordinates (latitude and longitude). Around this point, a square-shaped Area of Interest (AOI) of predetermined size is generated and divided into a regular grid with a predefined step, calculated in meters on the local projection plane. Each node of the grid represents a potential observation point. For each node, a spatial search algorithm queries the Google Street View API (Google Maps Platform, n.d.) to identify the nearest point that actually contains panoramic imagery, accepting the sample if the minimum distance is less than half of the previously defined step. This threshold ensures homogeneous distribution and prevents overlap between adjacent points. All valid points are stored in a CSV file structured according to standardised fields, including a unique identifier, geographic coordinates, orientation, acquisition date, and a validity index associated with spatial matching. This database constitutes the reference matrix for the subsequent visual acquisition phase, carried out automatically through Google Street View Static API (Google, n.d.). For each selected point, four images corresponding to the cardinal directions are extracted, each captured with a 90° field of view and a resolution of 640×640 pixels. The four projections generated for each node are then composited into a single synthetic panoramic view, obtained through the sequential rotation and merging of the cardinal images. The views thus obtained constitute a homogeneous georeferenced visual corpus capable of systematically representing the perceptual complexity of the territorial context and serving as the empirical basis for the subsequent phases of automatic classification and simulated cognitive evaluation.

The dataset provides the foundation for two complementary but distinct analytical workflows: the first, classificatory in nature, processes the images to identify homogeneous territorial classes; the second, evaluative in nature, analyzes the judgments and texts generated by the agents to infer perceptual polarity and gradients of landscape quality.

In the first phase, dedicated to automatic classification, each image is processed by LLaVA (Liu et al., 2024), an image-to-text AI model that, following specific instructions provided through prompts, generates a concise and neutral description, translating into language the salient visual features, the morphology of buildings, the spatial layout, the presence of natural elements, and the relationships of continuity or threshold. The dataset, composed of the textual descriptions produced by LLaVA for each image, is divided into batches of 100 records to ensure the computational feasibility of local processing. Each batch is then processed by the DeepSeek-9B model (DeepSeek-AI, 2025), which, through a specifically designed prompt, generates between five and ten synthetic macro-topics summarizing the main semantic trends present in the group of descriptions. The output consists of an unnumbered list of general themes formulated by the model as linguistic categories, without direct references to the source text. The results produced in the different iterations are subsequently subjected to a semantic deduplication process carried out using the same model, through a second prompt instructing the system to eliminate redundancies, unify synonyms and lexical variants, and produce synthetic and coherent categories capable of representing the entire semantic domain. Once the semantic structure has been consolidated and the emerging territorial classes defined, the process returns to the visual dimension. LLaVA is employed again for the automatic assignment of categories to images. Each frame in the dataset is provided with the list of classes identified during the clustering phase, and the model is instructed, through a prompt, to select the one most consistent with the visual content of the image. The assignment is accompanied by a percentage value expressing the degree of reliability of the recognition, calculated based on the correspondence between the generated description and the proposed category. This step makes it possible to validate the semantic–visual consistency of the entire system, reducing interpretive ambiguities and ensuring internal control over the quality of the clustering. The classes thus determined do not derive from predefined theoretical taxonomies, as in traditional land-use or zoning systems, but instead emerge from the data themselves through a process of semantic self-organisation that translates the visual continuity of the territory into an interpretable informational structure. From this perspective, the language generated by the models becomes an additional level of representation, a cognitive interface between image and territorial knowledge, capable of describing in a distributed and non-hierarchical way the spatial relationships that contribute to defining the form and perception of the landscape.

In parallel, the same images are subjected to a simulated cognitive evaluation through an LLM-as-Judge framework involving four multimodal agents configured as virtual observers: the Resident, focused on everyday usability and perceived care; the Tourist, who assesses the identity and overall recognizability of the place; the Architect, attentive to formal coherence, morphological clarity, and compositional quality; and the Engineer, concerned with continuity, accessibility, and maintenance conditions (Table 1). Each agent interprets the images according to its perceptual profile and produces, for each view, a numerical score on a scale from one to five along with a brief textual comment. The evaluation criteria are shared across agents and concern the overall quality of the place, formal coherence, pedestrian continuity and accessibility, spatial legibility, and maintenance condition.

Agent Profile	Prompt configuration
Resident (Care)	You are a resident observer assessing everyday usability, proximate usable greenery, opportunities for informal rest, and perceived care of frontages and edges
Tourist (Identity)	You are a cultural visitor (tourist) evaluating overall place quality, distinctiveness/identity when visible, and readability of the public realm configuration.
Academic Architect (Rigor)	You are an academic architect assessing formal coherence of street walls, material and chromatic consistency, and morphological clarity of open-space structure.
Construction Engineer (Feasibility)	You are a civil engineer assessing pedestrian continuity and barrier-free access, edge logic and thresholds, and observable maintenance/condition of built elements.

Table 1. Profiles of the simulated evaluators (simulated agents).

The result is a parallel dataset of simulated judgments, in which scores and verbal narratives combine to construct a cognitive geography of the landscape, representing perception not as a subjective event but as a comparable form of data. The texts generated by the simulated evaluators are subsequently analyzed using the VADER model (Hutto & Gilbert, 2014), which detects linguistic polarity and determines the positive, neutral, or negative valence of the judgments. The integration of numerical scores and textual polarities makes it possible to highlight patterns of coherence, perceptual divergences, and gradients of perceived quality. All data are then unified into a single georeferenced dataset that includes, for each point, coordinates, images, linguistic descriptions, territorial classes, scores, judgments, and sentiment evaluations.

This method enables the combination of morphological analysis, visual classification, and linguistic evaluation within an integrated system for the automated reading of the landscape, capable of producing a computational representation of the territory that brings together the objective dimension of data and the interpretive dimension of perception.

3. Case study and results

To test the methodological framework described above, the automated analysis process was applied in real urban context sufficiently articulated to reveal the morphological and perceptual variety of the contemporary landscape. The selected area is in Italy, in the town center of the Umbrian city of San Giustino (Figure 2), with its epicenter in Palazzo Bufalini, a historical and symbolic node around which various forms of urbanization and rural edges are organized. The area chosen for the analysis is a square of 1 km² around the palace, thus including both the central urban core and the immediately adjacent suburban and rural zones according to traditional zoning of the municipal master plan.

To map and characterize this heterogeneous space, the systematic sampling method described above was applied. The area was divided using a grid with a 40-meter step, both in the north–south and east–west directions, resulting in a network of observation points where a street-level panoramic image was acquired (Figure 3), composed of views in the four cardinal directions 0°, 90°, 180°, 270° assembled into a collage (Figure 4). This operation produced a georeferenced visual dataset of several hundred scenes, specifically 314 images, providing comprehensive coverage of the entire area surrounding Bufalini Palace.

Figure 2. Aerial view of the urban center of San Giustino centered in Palazzo Bufalini (© Google Earth 2025).

Through AI-assisted analysis, each panoramic image was automatically classified according to the prevailing landscape type, producing a visual zoning of the urban composition of San Giustino. The resulting map (Figure 5) highlights six recurring territorial categories (Table 2) that describe the perceptual and morphological continuity from the compact historical core to the surrounding rural areas.

The most extensive category, Suburban housing with gardens (40.81%), corresponds to the low-density residential fabric that develops around the historic center. These areas are characterized by single-family houses, widespread private greenery, and a general sense of spatial openness, where the relationship between built form and ground produces an ordered yet fragmented perception.

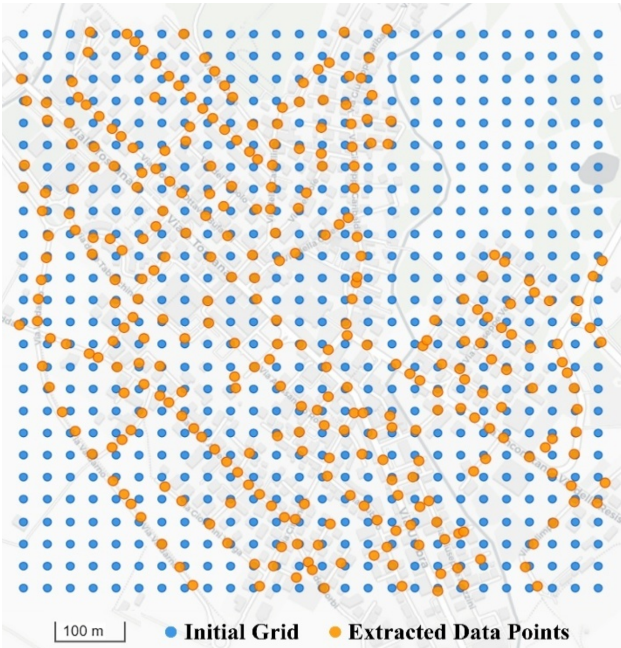

Figure 3. Map of the urban area showing the superimposed initial sampling grid and the points selected for image extraction from Google Street View API. The blue points represent the initial grid, while the orange points indicate the locations from which images were extracted.

Figure 4. Example of the synthetic panoramic view generated by merging the four cardinal images (0°, 90°, 180°, 270°) for the same observation point.

Category	Count	Percentage
Suburban housing with gardens	131	40.81 %
Urban fringes	90	28.04 %
Dense residential areas	33	10.28 %
Scattered rural housing	24	7.48 %
Rural/agricultural areas	20	6.23 %
Urban green spaces	16	4.98 %

Table 2. Distribution of landscape categories on AI-based classification.

This condition, typical of twentieth-century Umbrian suburban expansions, represents the contemporary form of the threshold between the urban and rural realms. The second most prevalent category, Urban fringes (28.04%), includes edge areas that are heterogeneous and constantly changing, where built, productive, and open spaces interweave without a defined hierarchy. These places, marked by morphological and functional discontinuities, express the suspended and unstable condition typical of peri-urban landscapes. Dense residential areas (10.28%) are concentrated in the historic core and along the main urban corridors. The continuity of facades, consistent height, and scarcity of open spaces define a strongly anthropized environment perceived as unified and compact. Here, building density translates into a visual saturation that constitutes the identity reference of the urban landscape. At the opposite end of the gradient lie Rural/agricultural areas (6.23%) and Scattered rural housing (7.48%), representing the persistence of the productive landscape and its gradual hybridization with dispersed built forms. In these areas, the regularity of the agricultural pattern and the sporadic presence of rural structures create a more homogeneous and stable perception, where natural and anthropic components are balanced in perceptual continuity. Finally, Urban green spaces (4.98%) represent a quantitatively limited but qualitatively significant presence, introducing rhythmic pauses within the built sequence. Parks and public gardens play a regulatory role, restoring the legibility of the urban fabric and enhancing ecological connectivity among its different parts. The classification outlines a continuous gradient from built to rural, in which the traditional dichotomy between city and countryside dissolves into a perceptual continuum. The findings are confirmed by the municipal development plan, which lends formal validity to the study’s conclusions.

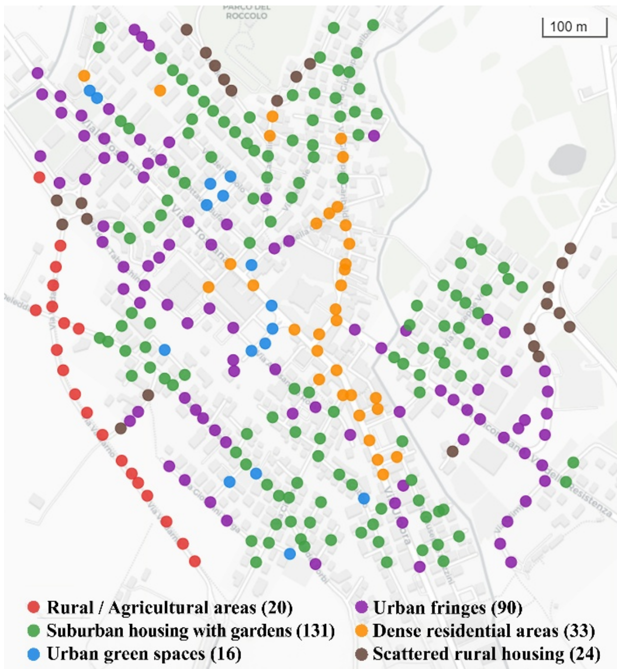

Figure 5. Map of the landscape classes identified through AI-assisted analysis. The distribution highlights the predominance of low-density residential areas with private gardens and the presence of transitional urban fringes towards rural zones.

Next, the simulated cognitive evaluation, conducted through the four multimodal agent profiles, Resident, Architect, Engineer, and Tourist, provides a comparative reading of the urban landscape based on different perceptual and disciplinary perspectives (Table 3).
The Resident displays the widest and most regular curve, with higher average values in Place Quality (3.30) and Maintenance State (2.95). This is a pragmatic, everyday gaze, sensitive to usability and the perception of care within the context, more inclined to recognize comfort and continuity even in areas of modest formal quality. The Architect, by contrast, shows greater selectivity, with lower scores in Place Quality (2.18) but higher ones in Formal Coherence (2.95) and Space Legibility (3.34), confirming a critical and disciplinary approach that privileges compositional clarity and morphological legibility. The Engineer occupies an intermediate position, assigning high values to Maintenance State (3.03) and Pedestrian Access (2.64), interpreting quality in terms of efficiency and functional continuity rather than aesthetics. Finally, the Tourist shows a regular but slightly lower distribution (2.43–2.80), interpreting the landscape in perceptual and symbolic terms, with a preference for recognizability and visual impact.

Agent	Place Quality	Formal Coherence	Pedastr. Access	Space Leg.	Maint. State
Resident	3.30	2.72	2.54	2.95	2.95
Architect	2.18	2.95	2.60	3.34	2.88
Engineer	2.12	2.81	2.64	3.13	3.03
Tourist	2.50	2.43	2.59	2.76	2.80

Table 3. Mean evaluation for each simulated profile.

These differences, though numerically modest, describe a coherent system of perceptual priorities: experiential for the Resident, critical for the Architect, functional for the Engineer, and empathetic for the Tourist. AI thus translates the plurality of viewpoints into a shared language, making comparable perceptions that, in reality, often remain invisible.
The analysis by territorial classes (Figure 6) further explores the relationship between judgment and morphology. In dense and compact contexts (Dense residential areas), the Architect and Engineer reward formal coherence and spatial legibility, while the Resident reports reduced comfort due to the scarcity of open and green spaces. Suburban areas with gardens show the highest convergence among profiles, with uniformly positive scores in Place Quality and Maintenance State: these are the places of balanced everyday life, where comfort and compositional order coincide. In Urban fringes, transitional and discontinuous zones, the greatest divergence emerges: technical evaluators penalize Formal Coherence, while the Resident and Tourist maintain more stable values, suggesting a perception linked to livability rather than structure. Urban green spaces invert the hierarchy: the Tourist assigns the highest values in Place Quality, recognizing greenery as an identity and regenerative element, while the other profiles maintain more neutral evaluations. In rural contexts and scattered housing areas, the curves flatten: morphological simplicity produces a shared perception of stability and legibility, where spatial linearity translates into visual calm.
Overall, the combined reading of the two levels (cognitive profiles and territorial classes) highlights how perceived quality emerges not from an absolute parameter but from the interaction among form, function, and perceptual sensitivity. Artificial intelligence, through multi-agent simulation, does not replace human vision; it extends it, translating the complexity of a collective gaze on the landscape into comparable data.

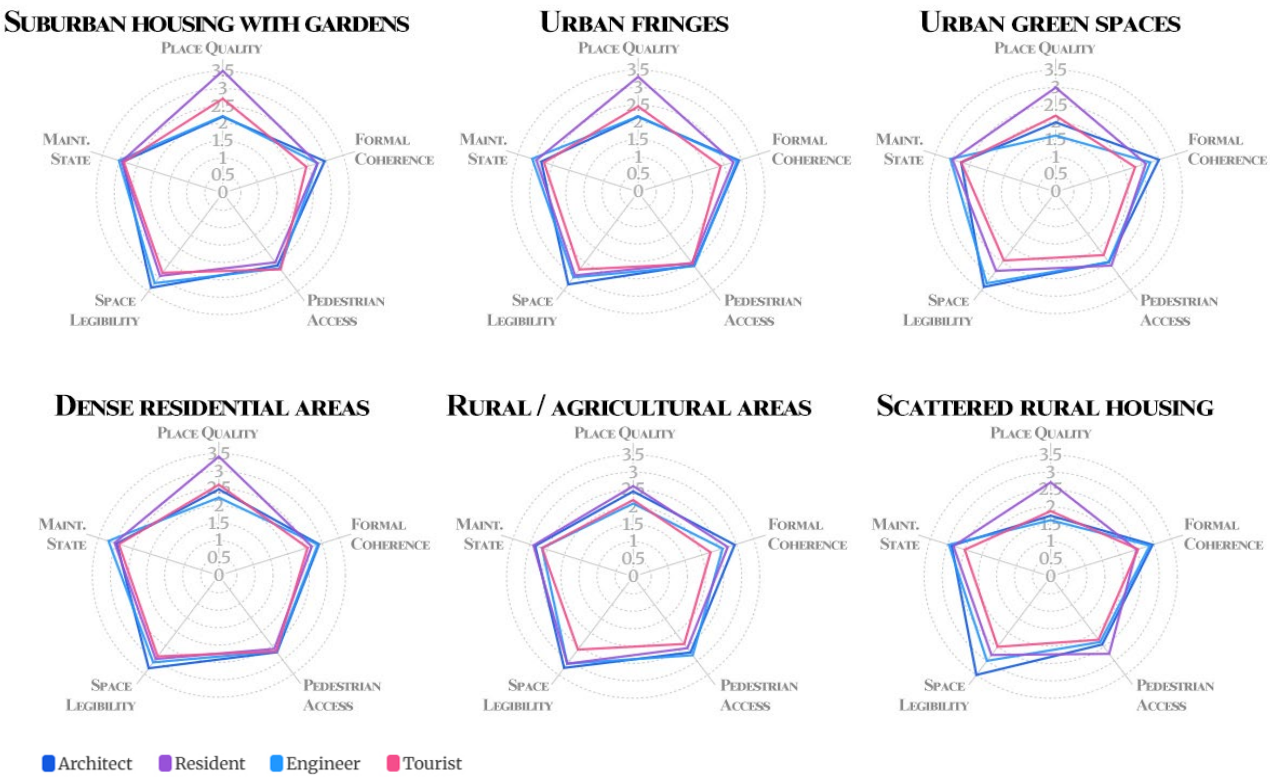

Figure 6. Mean evaluation of spatial quality indicators by agent across landscape categories.

This cognitive representation forms the basis for the subsequent spatial analysis, in which the judgments are projected onto the map to construct a perceptual geography of the territory. The cartographic transposition of the results (Figure 7) makes it possible to move from the descriptive level of the evaluations to the spatial level of representation.

The integration of numerical scores and linguistic polarity, analyzed through the VADER model, generates a cognitive map of the landscape in which images are translated into georeferenced evaluative points. Each point synthesizes the data derived from the pipeline, including territorial class, average scores from the four agents, and sentiment value, providing a distributed reading of perceived quality. To support the reliability of the method, a sample-based validation was conducted through independent evaluations by human observers. A subset of images was assessed according to the same criteria adopted in the multi-agent framework; the comparison revealed a high degree of correspondence both in score distribution and in overall interpretation, confirming the consistency of the automated process.

The sentiment map shows a predominance of neutral and slightly positive tones, with a range oscillating between –0.23 and +0.38. Positive values are concentrated in the central areas and along the main crossing axes, where morphological continuity and the presence of urban greenery contribute to a more stable and coherent perception. Conversely, edge zones, particularly urban fringes and some dense urban segments, display an increase in negative polarities, corresponding to visual discontinuities, residual spaces, and low perceptual legibility. The chromatic variation thus reflects not a simple aesthetic gradient but the measure of a shared sensitivity: the evaluators’ emotional responses synthesize the relational quality between built form and landscape, translating perception into territorial information.

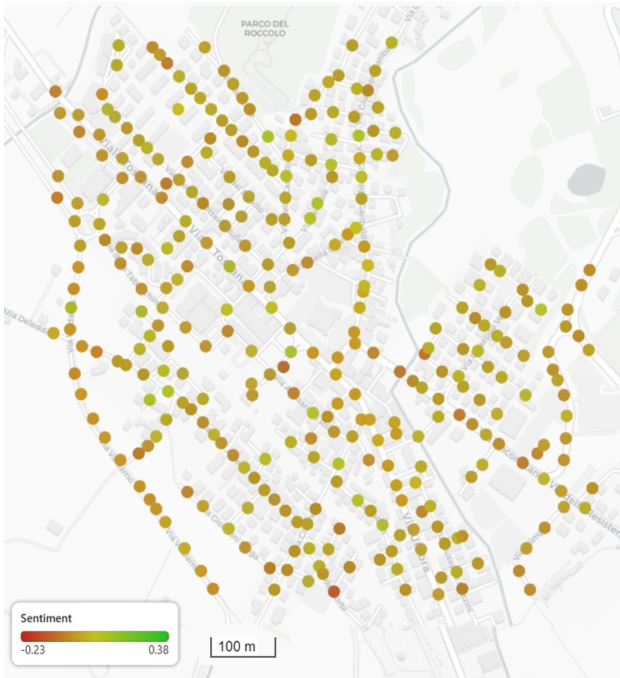

Figure 7. Sentiment map based on mean value of the evaluations of the four simulated profile: spatial distribution of perceived quality values, with color gradients indicating positive (green) and negative (red) emotional responses.

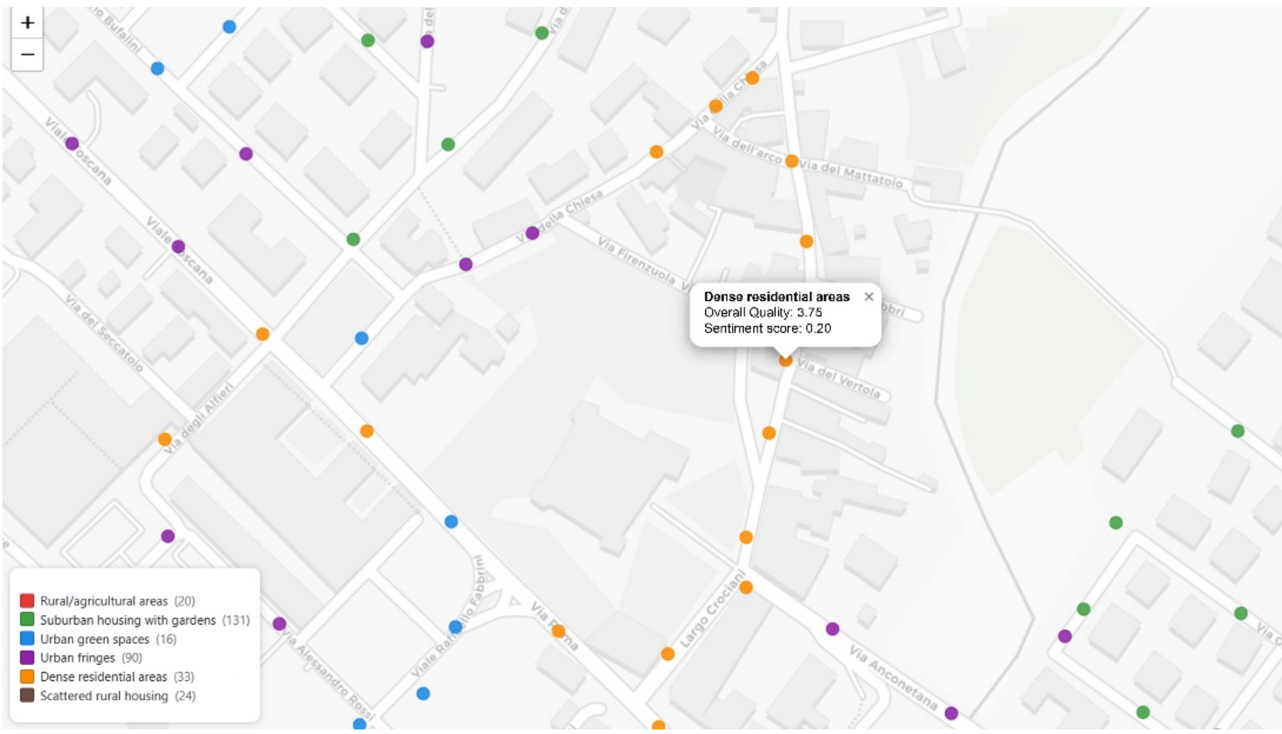

Figure 8. Interactive map of evaluated points: automatic classification of urban and rural areas based on the prevailing landscape type, with average overall quality and mean sentiment scores.

The multi-agent synthesis map (Figure 8) combines qualitative and quantitative information, representing for each grid node the territorial class, average score, and associated polarity. The result is a high-resolution interpretive system capable of revealing patterns consistent with the urban structure: suburban residential areas show high and homogeneous values, dense cores display moderate but stable perceived quality, while transitional areas reveal the greatest variability, indicating a discontinuous and unstable perception. In this representation, the landscape no longer appears as a sum of forms but as a stratified cognitive surface, where each judgment becomes a point within a field of perceptions and the subjective dimension is rendered objective through number and repetition.

The resulting cartography thus constitutes an expanded representation of reality, in which visual, linguistic, and evaluative data converge into a computational synthesis.

4. Conclusion

The experiments described reveal how the multimodal approach adopted, makes it possible to translate street-level imagery into an actual cognitive cartography in which visual, linguistic, and evaluative dimensions converge into a unified representation. The automatic zoning derived from the model-generated descriptions provides a stable morphological structure; the multi-agent evaluation projects onto this structure a plural yet comparable perception of quality; the sentiment analysis reveals its tonal dimension, rendering the landscape not only as form but as a field of measurable and mappable perceptions.

Three main results emerge from this process. On the methodological level, the workflow integrates description, semantic classification, and simulated evaluation within a coherent and reproducible pipeline capable of producing georeferenced data that can be interpreted and reused in analytical and design processes. On the epistemic level, representation is redefined as an inference across media, from pixel to text, from text to judgment, in which artificial intelligence extends the designer’s gaze without replacing it, making explicit and verifiable the evaluative criteria that are usually implicit. On the operational level, the results reveal interpretable spatial patterns: perceptual convergence in suburban fabrics, stability in dense centers, and discontinuity in fringe areas. These configurations constitute useful indicators for guiding strategies of planning, maintenance, and landscape enhancement.

Looking ahead, the same framework could evolve into a dynamic decision-making interface capable of linking automated analysis and design, allowing users to query classes, filter results by agent profile, and simulate alternative scenarios by evaluating their perceptual impact. In this vision, artificial intelligence is not a black box but a transparent interlocutor, a device that makes explicit the relationships between spatial structure and perceived quality, transforming representation into an act of shared knowledge. Consequently, representing the landscape under AI evaluation does not mean automating judgment, but rather restoring to the figure of the designer, understood as an algorithmic curator, the role of guiding the indeterminacy of generative processes toward forms of conscious knowledge, where the image becomes a critical synthesis between data and perception, between measurement and interpretation.

References

Antrop, M., 2005. Why landscapes of the past are important for the future. *Landscape and Urban Planning*, 70(1–2), 21–34. doi.org/10.1016/j.landurbplan.2003.10.002.

Berque, A., 1995. *Les raisons du paysage: de la Chine antique aux environnements de synthèse*. Hazan, Paris.

Bianconi, F., Filippucci, M., 2019. *Landscape Lab. Drawing, Perception and Design for the Next Landscape Models*, Vol. 20. Springer Nature.

Bianconi, F., Filippucci, M. (Eds.), 2021. *Digital Draw Connections*, Vol. 107. Springer International Publishing. doi.org/10.1007/978-3-030-59743-6.

Bianconi, F., Filippucci, M., Migliosi, A., Mommi, C., 2024. AI-D. Algorithmic experiments with AI to aid drawing. *SCIRES-IT*, 14(2), 31–48. doi.org/10.2423/i22394303v14n2p31.

Bommasani, R., 2021. On the opportunities and risks of foundation models. arXiv preprint. doi.org/10.48550/arXiv.2108.07258.

Carpo, M., 2017. The Second Digital Turn. MIT Press, Cambridge. doi.org/10.7551/mitpress/9976.003.0004.

Carpo, M., 2023. *Beyond digital: design and automation at the end of modernity*. MIT Press, Cambridge.

Corboz, A., 1985. Il territorio come palinsesto. *Casabella*, 516, 22–27.

Corner, J., 1999. The agency of mapping: Speculation, critique and invention. In: Cosgrove, D. (Ed.), *Mappings*. Reaktion Books, London, 213–252.

Cosgrove, D., 1984.*Social Formation and Symbolic Landscape*. University of Wisconsin Press, Madison.

Cosgrove, D. E., 2008. *Geography and Vision: Seeing, Imagining and Representing the World*. I.B. Tauris.

de Rubertis, R., 1994. *Il Disegno dell’Architettura*. NIS, Roma.

DeepSeek-AI, 2025. DeepSeek-R1: Incentivizing reasoning capability in LLMs via reinforcement learning. arXiv preprint. doi.org/10.48550/arXiv.2501.12948.

Floridi, L., 2019. *The Logic of Information: A Theory of Philosophy as Conceptual Design*. Oxford University Press, Oxford.

Google Maps Platform, n.d. Street View API. developers.google.com/maps/documentation/streetview.

Hutto, C., Gilbert, E., 2014. VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proc. Int. AAAI Conf. Web Soc. Media*, 8(1), 216–225. doi.org/10.1609/icwsm.v8i1.14550.

Liu, H., Li, C., Li, Y., Li, B., Zhang, Y., Shen, S., Lee, Y. J., 2024. LLaVA-NeXT: Improved reasoning, OCR, and world knowledge. llava-vl.github.io/blog/2024-01-30-llava-next/.

Lynch, K., 1960. *The Image of the City*. MIT Press, Cambridge.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B. A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.*, 152, 166–177. doi.org/10.1016/j.isprsjprs.2019.04.015.

Mai, G., Huang, W., Sun, J., Song, S., Mishra, D., Liu, N., Gao, S., Liu, T., Cong, G., Hu, Y., 2024. On the opportunities and challenges of foundation models for GeoAI. *ACM Trans. Spatial Algorithms Syst.*, 10(2), 1–46.

Manovich, L., 2020. *Cultural Analytics*. MIT Press. doi.org/10.7551/mitpress/11214.001.0001.

Mitchell, W. J. T., 2002. *Landscape and Power*. University of Chicago Press, Chicago.

Pallasmaa, J., 1994. Identity, intimacy and domicile. Notes on the phenomenology of home. *Finnish Architectural Review*, 1–16.

Peng, Z., Lu, W., An, H., Xia, X., Zhang, Y., Xue, F., Chen, J., 2025. Vision-language model (VLM)-enabled street view analytics: A systematic literature review. *Engineering, Construction and Architectural Management*, 1–19.

Ugo, V., 1989. Rappresentare/Costruire. *XY: Le Dimensioni del Disegno*, 10.

Wei, J., Yue, W., Li, M., Gao, J., 2022. Mapping human perception of urban landscape from street-view images: A deep-learning approach. *Landscape and Urban Planning*, 218, 104286. doi.org/10.1016/j.landurbplan.2021.104286.

Wu, J., 2013. Landscape sustainability science: ecosystem services and human well-being in changing landscapes. *Landscape Ecology*, 28(6), 999–1023. doi.org/10.1007/s10980-013-9894-9.

Xing, Y., Gan, W., Chen, Q., 2025. Artificial intelligence in landscape architecture: A survey. *Int. J. Machine Learning and Cybernetics*, 16(7), 4685–4710.

Zhang, J., Li, Y., Fukuda, T., Wang, B., 2025. Urban safety perception assessments via integrating multimodal large language models with street view images. *Cities*, 165, 106122.

Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., Torralba, A., 2018. Places: A 10 million image database for scene recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 40(6), 1452–1464. doi.org/10.1109/TPAMI.2017.2723009.